

Robust Optimization for Nonlinear Dynamic Systems

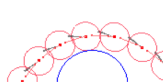
Moritz Diehl

Systems Control and Optimization Laboratory
Faculty of Engineering & Faculty of Mathematics and Physics
University of Freiburg, Germany

based on joint work with
Boris Houska (ShanghaiTech),
Joris Gillis (KU Leuven) and Greg Horn (google)

- Robust optimization problem statement
- Two conservative approximation approaches
 - Linearization
 - Lagrangian relaxation
- Dynamic problem statement, and solution by
 - forward sensitivities
 - adjoint sensitivities
 - Lyapunov matrix propagation

Problem Statement



$$\begin{array}{ll} \underset{u \in \mathbb{R}^{n_u}}{\text{minimize}} & \max_{\|w\|_2 \leq 1} F_0(u, w) \\ \text{subject to} & \max_{\|w\|_2 \leq 1} F_i(u, w) \leq 0, \quad i = 1, \dots, n_F. \end{array}$$

- relevant dimensions: n_u, n_w, n_F
- interested in case $n_w \gg 1$ (making sampling expensive)
- game interpretation: we choose $u \in \mathbb{R}^{n_u}$, then adverse player (nature) chooses $w \in \mathbb{R}^{n_w}$
- unit ball can represent all ellipsoidal uncertainties (can generalize to other sets described by inequalities)

Problem Statement: Compact Formulation

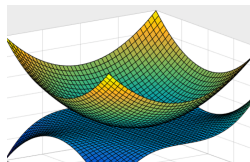


$$\begin{array}{ll} \underset{u \in \mathbb{R}^{n_u}}{\text{minimize}} & F_0^{\text{exac}}(u) \\ \text{subject to} & F_i^{\text{exac}}(u) \leq 0, \quad i = 1, \dots, n_F. \end{array}$$

with $F_i^{\text{exac}}(u) := \max_{\|w\|_2 \leq 1} F_i(u, w)$.

- Aim 1: find computationally tractable conservative approximations for $F_i^{\text{exac}}(u)$ (i.e. tight upper bounds)
- Aim 2: solve overall problem to local optimality w.r.t. u with structure exploiting nonlinear programming (NLP) method

Assumption throughout the talk: bounded 2nd derivatives



$$\begin{aligned} & \underset{u \in \mathbb{R}^{n_u}}{\text{minimize}} && \max_{\|w\|_2 \leq 1} F_0(u, w) \\ & \text{subject to} && \max_{\|w\|_2 \leq 1} F_i(u, w) \leq 0, \\ & && i = 1, \dots, n_F \end{aligned}$$

ASSUMPTION

There exist positive smooth functions $L_i(u)$ such that for all $w \in B := \{w \in \mathbb{R}^{n_w} \mid w^\top w \leq 1\}$ holds:

$$\nabla_w^2 F_i(u, w) \preceq L_i(u) I$$

Bounds the non-concavity of F_i w.r.t. w .

First Approach: Approximation by Linearization

Using Taylor's theorem, for each $w \in B$ there exists a $t \in [0, 1]$ such that

$$F_i(u, w) = F_i(u, 0) + \nabla_w F_i(u, 0)^\top w + \frac{1}{2} \underbrace{w^\top \nabla_w^2 F_i(u, tw) w}_{\leq L_i(u)}.$$

Yields upper bound (using self duality of the Euclidean norm)

$$\underbrace{\max_{w \in B} F_i(u, w)}_{=: F_i^{\text{exac}}(u)} \leq \underbrace{F_i(u, 0) + \|\nabla_w F_i(u, 0)\|_2 + \frac{1}{2} L_i(u)}_{=: F_i^{\text{lin}}(u)}$$

[Nagy & Braatz, JPC, 2004]

Approximation by Linearization (Conservative)

$$\begin{aligned} & \underset{u \in \mathbb{R}^{n_u}}{\text{minimize}} && F_0(u, 0) + \|\nabla_w F_0(u, 0)\|_2 + \frac{1}{2}L_0(u) \\ & \text{subject to} && F_i(u, 0) + \|\nabla_w F_i(u, 0)\|_2 + \frac{1}{2}L_i(u) \leq 0, \quad i = 1, \dots, n_F. \end{aligned}$$

- This is a nonlinear Second Order Cone Program (SOCP)
- Could be solved with Sequential Convex Programming (SCP) or plain NLP
- Exact Hessian method needs third order derivatives
- For dynamic systems, different ways to obtain $\nabla_w F_i(x, 0)$:
 - forward sensitivities [Nagy & Braatz, JPC, 2004]
 - adjoint sensitivities [D., Bock, Kostina, Math. Prog., 2006]
 - Lyapunov matrix propagation [Houska & D., CDC, 2009]

Second Approach: Lagrangian Relaxation

Lower level maximization problem:

$$F_i^{\text{exac}}(u) = \max_{w \in \mathbb{R}^{n_w}} F_i(u, w) \quad \text{s.t.} \quad \frac{1}{2}(w^\top w - 1) \leq 0$$

Lagrangian: $\mathcal{L}(u, w, \lambda) = F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1)$

Lagrange dual function:

$$d_i(u, \lambda) = \max_{w \in B} \left(F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1) \right)$$

Weak duality and relaxation gives upper bound:

$$F_i^{\text{exac}}(u) \leq \min_{\lambda \geq 0} d_i(u, \lambda)$$

Second Approach: Lagrangian Relaxation

Lower level maximization problem:

$$F_i^{\text{exac}}(u) = \max_{w \in \mathbb{R}^{n_w}} F_i(u, w) \quad \text{s.t.} \quad \frac{1}{2}(w^\top w - 1) \leq 0$$

Lagrangian: $\mathcal{L}(u, w, \lambda) = F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1)$

Lagrange dual function:

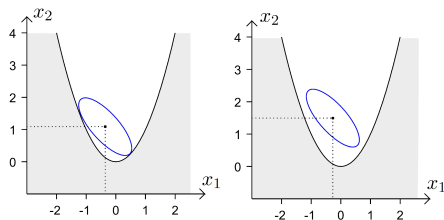
$$d_i(u, \lambda) = \max_{w \in B} \left(F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1) \right)$$

Weak duality and relaxation gives upper bound:

$$F_i^{\text{exac}}(u) \leq \min_{\lambda \geq 0} d_i(u, \lambda) \leq \min_{\lambda \geq L_i(u)} d_i(u, \lambda) =: F_i^{\text{lagr}}(u)$$

[Houska & D., Math. Prog. Ser. A, 2013]

How much conservatism is introduced?



THEOREM

Lagrangian relaxation is always tighter than linearization:

$$F_i^{\text{exac}}(u) \leq F_i^{\text{lagr}}(u) \leq F_i^{\text{lin}}(u)$$

If $F_i(u, w)$ is concave or quadratic in w , it is even exact.

[Yakubovich, Vestnik Leningrad Univ., 1971/1977]

[Houska & D., Math. Prog. Ser. A, 2013]

Lagrangian relaxation: convex lower level problems

Note that $(F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1))$ is concave in w for $\lambda \geq L_i(u)$: first order optimality conditions are sufficient

$$\begin{aligned} F_i^{\text{lagr}}(u) &= \min_{\lambda \geq L_i(u)} \max_{w \in B} \left(F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1) \right) \\ &= \min_{\lambda, w} \left(F_i(u, w) - \frac{\lambda}{2}(w^\top w - 1) \right) \\ &\text{s.t.} \quad \nabla_w F_i(u, w) - \lambda w = 0, \\ &\quad \lambda \geq L_i(u), \quad \|w\| \leq 1. \end{aligned}$$

- no complementarity condition needed
- constraint $\|w\| \leq 1$ deals with non-convexities outside B

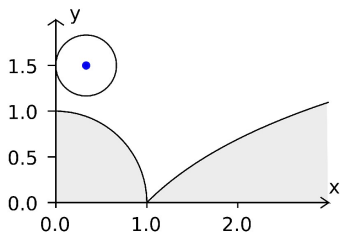
[D., Houska, Stein, Steuermann, *Comp. Opt. Appl.*, 2013]

Lagrangian relaxation based optimization problem

$$\begin{aligned} & \underset{u, \lambda_0, w_0, \dots, \lambda_{n_F}, w_{n_F}}{\text{minimize}} && \left(F_0(u, w_0) - \frac{\lambda_0}{2}(w_0^\top w_0 - 1) \right) \\ & \text{subject to} && \left(F_i(u, w_i) - \frac{\lambda_i}{2}(w_i^\top w_i - 1) \right) \leq 0, \\ & && i = 1, \dots, n_F. \\ & && \nabla_w F_j(u, w_j) - \lambda_j w_j = 0, \\ & && \lambda_j \geq L_j(u), \quad \|w_j\| \leq 1, \quad j = 0, 1, \dots, n_F. \end{aligned}$$

- equivalent to previous MPCC formulation, e.g. [Stein 2003]
- can solve with standard NLP, or Sequential Convex Bilevel Programming (SCBP) [Houska & D., Math. Prog. Ser. A, 2013]
- can show: no 3rd order derivatives needed for exact Hessian
- need $(n_F + 1)(n_w + 1)$ additional optimization variables

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

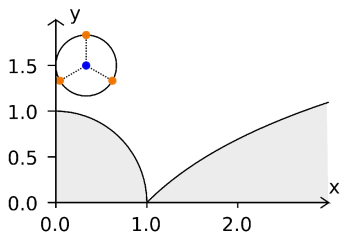
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

Ball with radius $r = \frac{1}{3}$, $B(v, w) := v^2 + w^2 - r^2 \leq 0$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

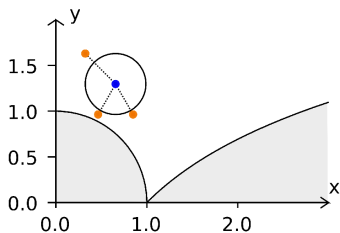
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

$$\text{Ball with radius } r = \frac{1}{3}, \quad B(v, w) := v^2 + w^2 - r^2 \leq 0$$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

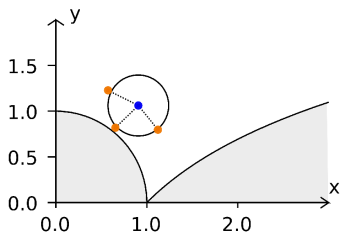
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

$$\text{Ball with radius } r = \frac{1}{3}, \quad B(v, w) := v^2 + w^2 - r^2 \leq 0$$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

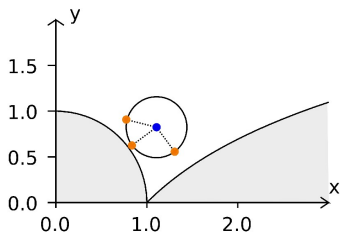
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

Ball with radius $r = \frac{1}{3}$, $B(v, w) := v^2 + w^2 - r^2 \leq 0$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

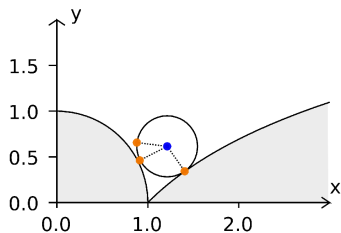
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

Ball with radius $r = \frac{1}{3}$, $B(v, w) := v^2 + w^2 - r^2 \leq 0$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

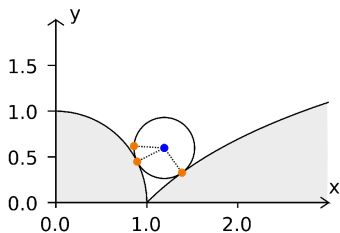
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

Ball with radius $r = \frac{1}{3}$, $B(v, w) := v^2 + w^2 - r^2 \leq 0$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

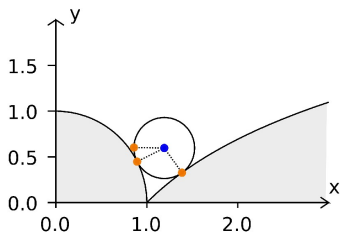
subject to

$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Uncertainty Set:

Ball with radius $r = \frac{1}{3}$, $B(v, w) := v^2 + w^2 - r^2 \leq 0$

Tutorial Example (2 uncertainties, 3 constraints)



$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2$$

subject to

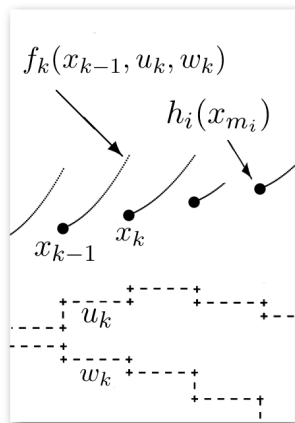
$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x + w)^2 - (y + v)^2 \\ 0 \geq \log(x + w) - (y + v) \end{cases}$$

Iteration	1	2	3	4	5	6	7	8
$-\log_{10}(\text{KKT-TOL})$	0.3	0.5	0.7	1.0	1.5	3.4	7.0	12.1

Can achieve high accuracy

- Robust optimization problem statement
- Two conservative approximation approaches
 - Linearization
 - Lagrangian relaxation
- **Dynamic problem statement, and solution by**
 - **forward sensitivities**
 - **adjoint sensitivities**
 - **Lyapunov matrix propagation**

Underlying Dynamic System



Now, the F_i are functions of states x_k that are generated by dynamic system:

$$x_k = f_k(x_{k-1}, u_k, w_k), \quad k = 1, \dots, N$$

with initial value

$$x_0 = f_0(u_0, w_0)$$

and inputs $u = (u_0, \dots, u_N)$ and $w = (w_0, \dots, w_N)$. Each $F_i = h_i(x_{m_i})$ is evaluated at one selected time point m_i .

- dimensions n_u , n_w or n_F often grow with horizon length N
- state dimension n_x can be smaller than n_w and n_F

Nominal Problem with Dynamic System

$$\begin{aligned} & \underset{u,x}{\text{minimize}} && h_0(x_{m_0}) \\ & \text{subject to} && h_i(x_{m_i}) \leq 0, \quad i = 1, \dots, n_F, \\ & && x_0 = f_0(u_0, w_0), \\ & && x_k = f_k(x_{k-1}, u_k, w_k), \quad k = 1, \dots, N. \end{aligned}$$

- for nominal problem, $w = (w_0, \dots, w_N)$ is fixed to zero
- robust problem easier to formulate using implicit representation of dynamic system (next slide)

Implicit Function Representation

Collecting all states in one vector $x = (x_0, \dots, x_N)$, one can summarize the dynamics by an implicit function

$$G(x, u, w) = 0, \quad H_i(x) := h_i(x_{m_i})$$

with

$$G(x, u, w) := \begin{bmatrix} f_0(u_0, w_0) - x_0 \\ f_1(x_0, u_1, w_1) - x_1 \\ f_2(x_1, u_2, w_2) - x_2 \\ \vdots \\ f_N(x_{N-1}, u_N, w_N) - x_N \end{bmatrix}.$$

Then, worst case functions are represented by

$$F_i^{\text{exac}}(u) = \max_{w, x} H_i(x) \quad \text{s.t.} \quad G(x, u, w) = 0, \quad w^\top w \leq 1.$$

Three ways to treat linearization based approximation

- forward sensitivities [Nagy & Braatz, JPC, 2004]
- adjoint sensitivities [D., Bock, Kostina, Math. Prog., 2006]
- Lyapunov matrix propagation [Houska & D. 2009, Gillis 2015]

Forward sensitivities

Represent $F_i(u, w)$ by $H_i(x)$ where x is implicitly defined by $G(x, u, w) = 0$. Get gradient by implicit function theorem,

$$\nabla_w F_i(u, 0)^\top = \nabla_x H_i(x)^\top \underbrace{\left(- \left(\frac{\partial G}{\partial x}(x, u, 0) \right)^{-1} \frac{\partial G}{\partial w}(x, u, 0) \right)}_{= \frac{dx}{dw} =: S}$$

Forward sensitivities use matrix variable S defined by

$$S = - \left(\frac{\partial G}{\partial x}(x, u, 0) \right)^{-1} \frac{\partial G}{\partial w}(x, u, 0)$$

or equivalently by

$$\frac{\partial G}{\partial w}(x, u, 0) + \left(\frac{\partial G}{\partial x}(x, u, 0) \right) \cdot S = 0.$$

Forward sensitivity problem statement

$$\begin{aligned} & \underset{u,x,S}{\text{minimize}} && H_0(x) + \|S^\top \nabla_x H_0(x)\|_2 + \frac{1}{2}L_0(u) \\ & \text{subject to} && H_i(x) + \|S^\top \nabla_x H_i(x)\|_2 + \frac{1}{2}L_i(u) \leq 0, \quad i = 1, \dots, n_F, \\ & && G(x, u, 0) = 0, \\ & && \frac{\partial G}{\partial w}(x, u, 0) + \left(\frac{\partial G}{\partial x}(x, u, 0) \right) \cdot S = 0. \end{aligned}$$

Need n_w extra variables of same dimension as x .

Very expensive for $n_w \gg 1$.

Adjoint sensitivities

Adjoint sensitivities divide the gradient expression differently

$$\nabla_w F_i(u, 0)^\top = \underbrace{-\nabla_x H_i(x)^\top \left(\frac{\partial G}{\partial x}(x, u, 0) \right)^{-1}}_{=:\lambda_i^\top} \frac{\partial G}{\partial w}(x, u, 0)$$

and introduce adjoint vector variables λ_i , $i = 0, \dots, n_F$ that are implicitly defined by

$$\nabla_x H_i(x) + \left(\frac{\partial G}{\partial x}(x, u, 0) \right)^\top \lambda_i = 0$$

Adjoint sensitivity problem statement

$$\begin{aligned} & \underset{u,x,\lambda}{\text{minimize}} && H_0(x) + \left\| \left(\frac{\partial G}{\partial w}(x, u, 0) \right)^\top \lambda_0 \right\|_2 + \frac{1}{2} L_0(u) \\ & \text{subject to} && H_i(x) + \left\| \left(\frac{\partial G}{\partial w}(x, u, 0) \right)^\top \lambda_i \right\|_2 + \frac{1}{2} L_i(u) \leq 0, \\ & && i = 1, \dots, n_F, \\ & && G(x, u, 0) = 0, \\ & && \nabla_x H_j(x) + \left(\frac{\partial G}{\partial x}(x, u, 0) \right)^\top \lambda_j = 0, \quad j = 0, 1, \dots, n_F. \end{aligned}$$

Note: compared to nominal problem, need n_F extra variables of the same dimension as x . Independent of noise dimension n_w .

Lyapunov Matrix Propagation

For linearization based approach

- forward sensitivities good for large n_F , small n_w
- adjoint sensitivities good for small n_F , large n_w

Dream:

a formulation that works well for large n_w and large n_F

A Close Look at the Forward Sensitivity Approach

$$\begin{aligned} & \underset{u, x, S}{\text{minimize}} && h_0(x_{m_0}) + \|S_{m_0}^\top \nabla h_0\|_2 + \frac{1}{2}L_0(u) \\ & \text{subject to} && h_i(x_{m_i}) + \|S_{m_i}^\top \nabla h_i\|_2 + \frac{1}{2}L_i(u) \leq 0, \quad i = 1, \dots, n_F, \\ & && x_0 = f_0(u_0, 0), \\ & && x_k = f_k(x_{k-1}, u_k, 0), \quad k = 1, \dots, N, \\ & && S_0 = \nabla_w f_0^\top, \\ & && S_k = \nabla_{x_{k-1}} f_k^\top S_{k-1} + \nabla_w f_k^\top, \quad k = 1, \dots, N. \end{aligned}$$

Have $(N + 1)$ matrices S_k of dimension $n_x \times n_w$, where n_x is the single stage state dimension. Expect $n_w = O(N)$, $n_w \gg n_x$.

Note that $\|S_{m_i}^\top \nabla h_i\|_2 = \sqrt{\nabla h_i^\top S_{m_i} S_{m_i}^\top \nabla h_i}$.

Replacing Sensitivities by Lyapunov Matrices

Regard $S_k = \nabla_{x_{k-1}} f_k^\top S_{k-1} + \nabla_w f_k^\top$. Define $A_k := \nabla_{x_{k-1}} f_k^\top \in \mathbb{R}^{n_x \times n_x}$ and

$$\nabla_w f_k^\top = (0 \cdots 0 | B_k | 0 \cdots 0) \in \mathbb{R}^{n_x \times n_w}$$

with $B_k := \nabla_w f_k^\top$. Also

$$S_{k-1} = (* \cdots * | 0 | 0 \cdots 0) \quad \text{and} \quad A_k S_{k-1} = (* \cdots * | 0 | 0 \cdots 0).$$

This implies $(A_k S_{k-1}) \nabla_w f_k = 0$, and thus

$$\begin{aligned} P_k &:= S_k S_k^\top &= (A_k S_{k-1} + \nabla_w f_k^\top)(A_k S_{k-1} + \nabla_w f_k^\top)^\top \\ & &= A_k S_{k-1} S_{k-1}^\top A_k^\top + \nabla_w f_k^\top \nabla_w f_k \\ & &= A_k S_{k-1} S_{k-1}^\top A_k^\top + B_k B_k^\top \\ & &= A_k P_{k-1} A_k^\top + B_k B_k^\top \end{aligned}$$

This is a Lyapunov matrix equation with dimension $P_k \in \mathbb{R}^{n_x \times n_x}$

Lyapunov Matrix Reformulation of Forward Sensitivities

$$\begin{aligned} \underset{u, x, P}{\text{minimize}} \quad & h_0(x_{m_0}) + \sqrt{\nabla h_0^\top P_{m_0} \nabla h_0} + \frac{1}{2} L_0(u) \\ \text{subject to} \quad & h_i(x_{m_i}) + \sqrt{\nabla h_i^\top P_{m_i} \nabla h_i} + \frac{1}{2} L_i(u) \leq 0, \\ & i = 1, \dots, n_F, \\ & x_0 = f_0(u_0, 0), \\ & x_k = f_k(x_{k-1}, u_k, 0), \quad k = 1, \dots, N, \\ & P_0 = B_0 B_0^\top, \\ & P_k = A_k P_{k-1} A_k^\top + B_k B_k^\top, \quad k = 1, \dots, N. \end{aligned}$$

Only need $(N + 1) \frac{n_x(n_x + 1)}{2}$ extra variables.
Independent of both n_w and n_F .

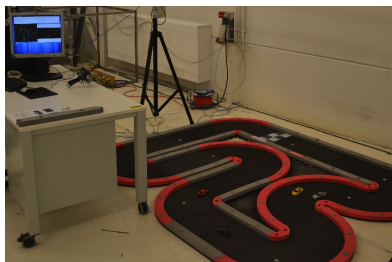
Infinite Time Horizons with Periodic Controls

Assuming stable periodic dynamics and zero bounds $L_i(u)$, can compute P_k resulting from infinite noise sequence w in ℓ_2 unit ball by Periodic Lyapunov Equation (PLE):

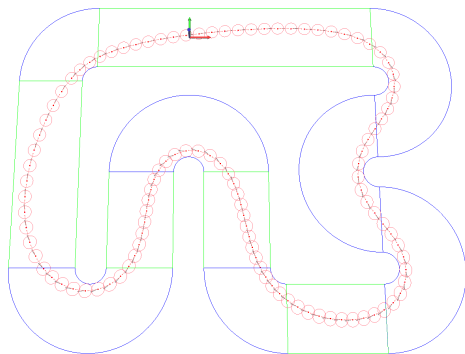
$$\begin{aligned} & \underset{u, x, P}{\text{minimize}} && h_0(x_{m_0}) + \sqrt{\nabla h_0^\top P_{m_0} \nabla h_0} \\ & \text{subject to} && h_i(x_{m_i}) + \sqrt{\nabla h_i^\top P_{m_i} \nabla h_i} \leq 0, \quad i = 1, \dots, n_F, \\ & && x_0 = x_N, \\ & && x_k = f_k(x_{k-1}, u_k, 0), \quad k = 1, \dots, N, \\ & && P_0 = P_N, \\ & && P_k = A_k P_{k-1} A_k^\top + B_k B_k^\top, \quad k = 1, \dots, N. \end{aligned}$$

Can treat PLE with periodic Schur decomposition [Varga 1997].
CPU time savings up to factor 100 possible [Gillis 2015].

Robust Race Cars (Greg Horn and Joris Gillis)



- 6 states, i.e. $n_x = 6$
- 100 time steps, i.e. $N = 100$
- 6 disturbances, i.e. $n_w = 600$
- 2 controls and 4 feedback gains, i.e. $n_u = 204$
- solved in 40 seconds using CasADi and IPOPT

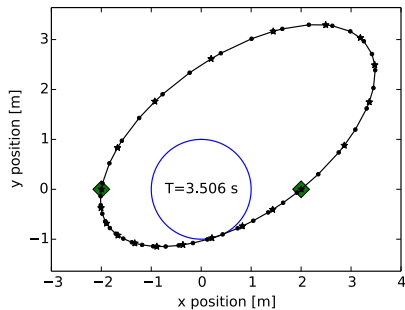


Race cars with online optimal control (Robin Verschueren)

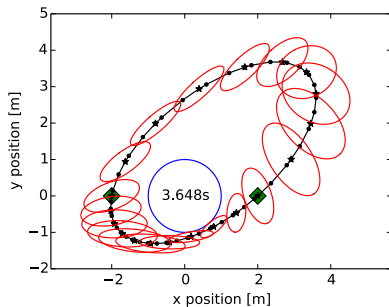
Quadcopter Flight Around Obstacle (Joris Gillis)



Nominal Solution



Robustified Solution



Conclusions

- Approximations by Linearization or Lagrangian relaxation lead to computationally tractable nonlinear programming problems
- Lagrangian relaxation is tighter but more expensive (avoids MPCC and needs only 2nd order derivatives)
- Linearization is cheaper and comes in three variants:
 - forward sensitivities: good for few uncertain parameters
 - adjoint sensitivities: good for few constraints
 - Lyapunov matrix propagation: good for small state dimensions
- In control applications, robust nonlinear dynamic optimization allows one to design trajectories and tune feedback controllers simultaneously